Высокодоступное приложение в кластере Kubernetes

Марсель Ибраев







Марсель Ибраев

CTO Слёрм

— Инженер с 8-летним стажем

— Certified Kubernetes Administrator

— Внедрения Kubernetes для клиентов Southbridge

— Разработчик курсов и спикер Слёрм

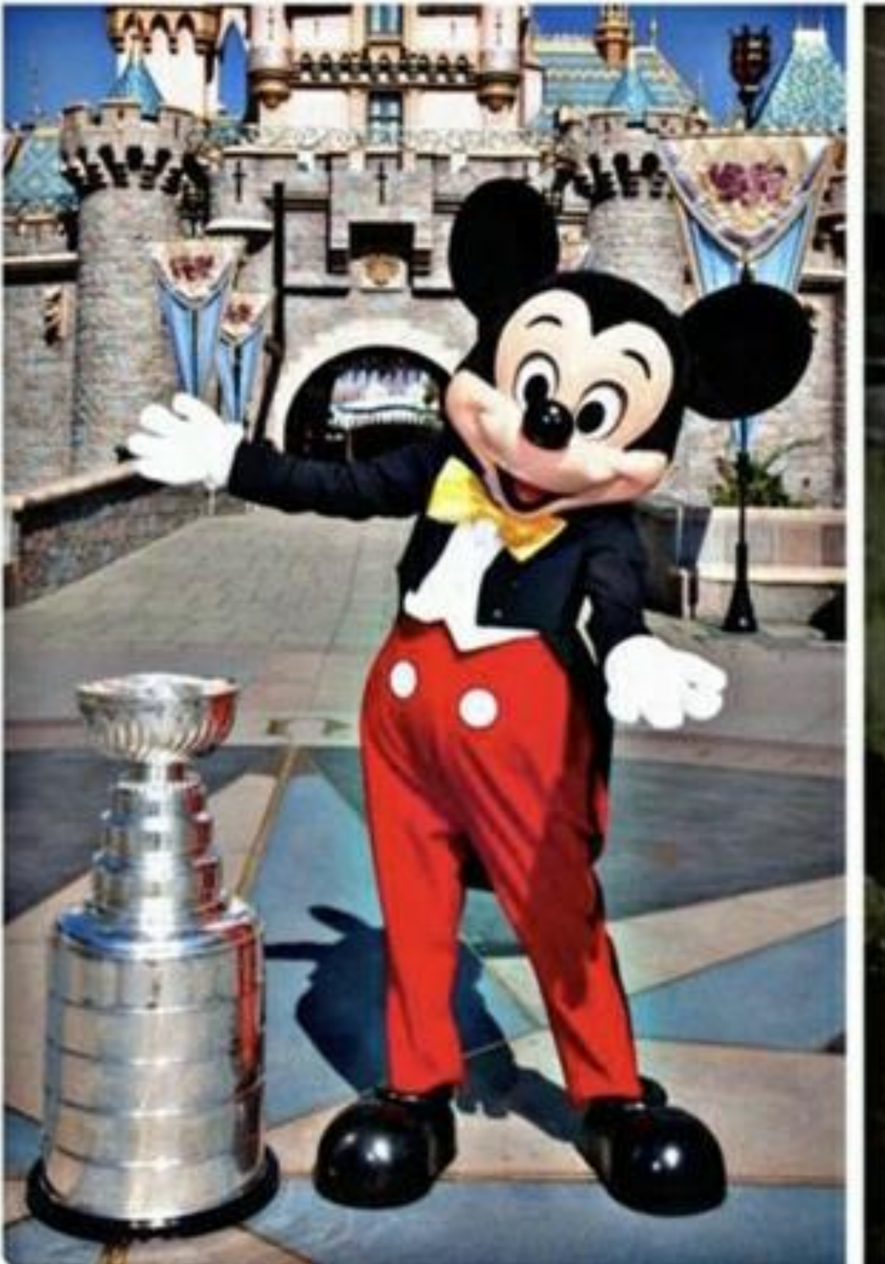



ПРИШЛА ПОРА
УМОПОМРАЧИТЕЛЬНЫХ СКАЗОВ

Регламентные работы:

Managed Kubernetes

Bare metal Kubernetes







История из жизни №1:

- Разработчики выкатывают релиз приложения;
- В это время админы перезагружают сервер;
- Единственная рабочая реплика приложения была на этом сервере;
- …
- PROFIT!



История из жизни №1:

- Разработчики выкатывают релиз приложения;
- В это время админы перезагружают сервер;
- Единственная рабочая репли а на этом сервере;
- …
- PROFIT!





Pod Disruption Budget

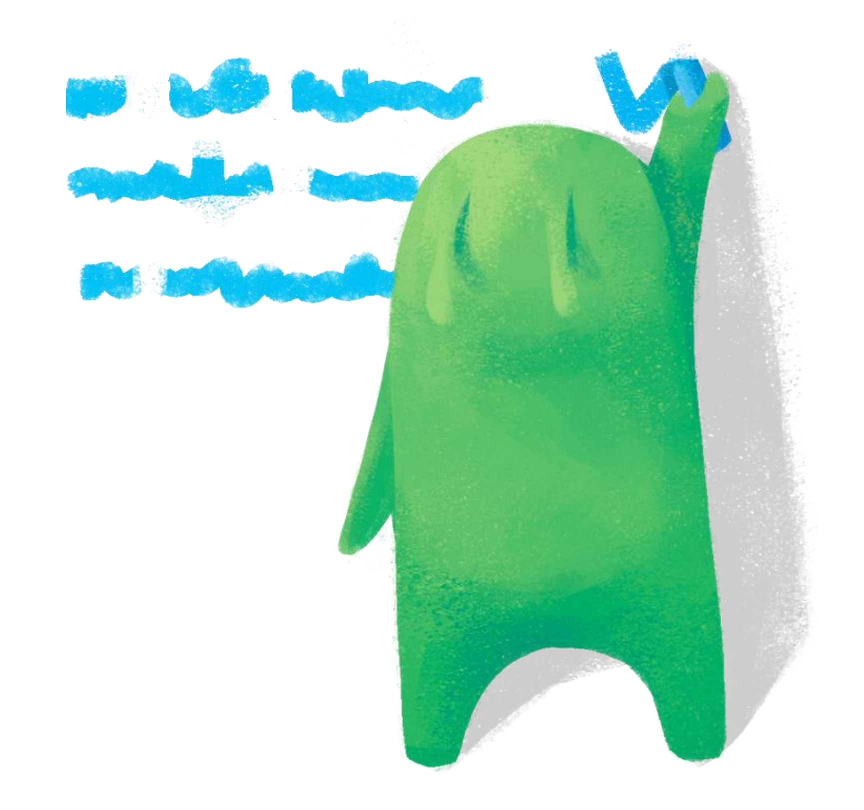
Ограничивает количество инстансов, которые могут быть недоступны одновременно

Администратор должен использовать “Eviction API”

https://kubernetes.io/docs/tasks/run-application/configure-pdb/



```yaml
apiVersion: policy/v1beta1
kind: PodDisruptionBudget
metadata:
  name: zk-pdb
spec:
  maxUnaviable: 1
  selector:
    matchLabels:
      app: zookeeper
```





{...} ingress.yaml	Temporary store kaniko 1.
{...} ipaddr-alertrule.yaml	Temporary store kaniko 1.
{...} job-review.yaml	Fix trailing spaces
{...} latency-alerts.yaml	Temporary store kaniko 1.
{...} pdb.yaml	Add PDB for production p
{...} secret-metrics-basicauth.yaml	Fix YAMLlint errors
{...} service-stand.yaml	Temporary store kaniko 1.
{...} service.yaml	Fix YAMLlint errors
{...} servicemonitor.yaml	Temporary store kaniko 1.

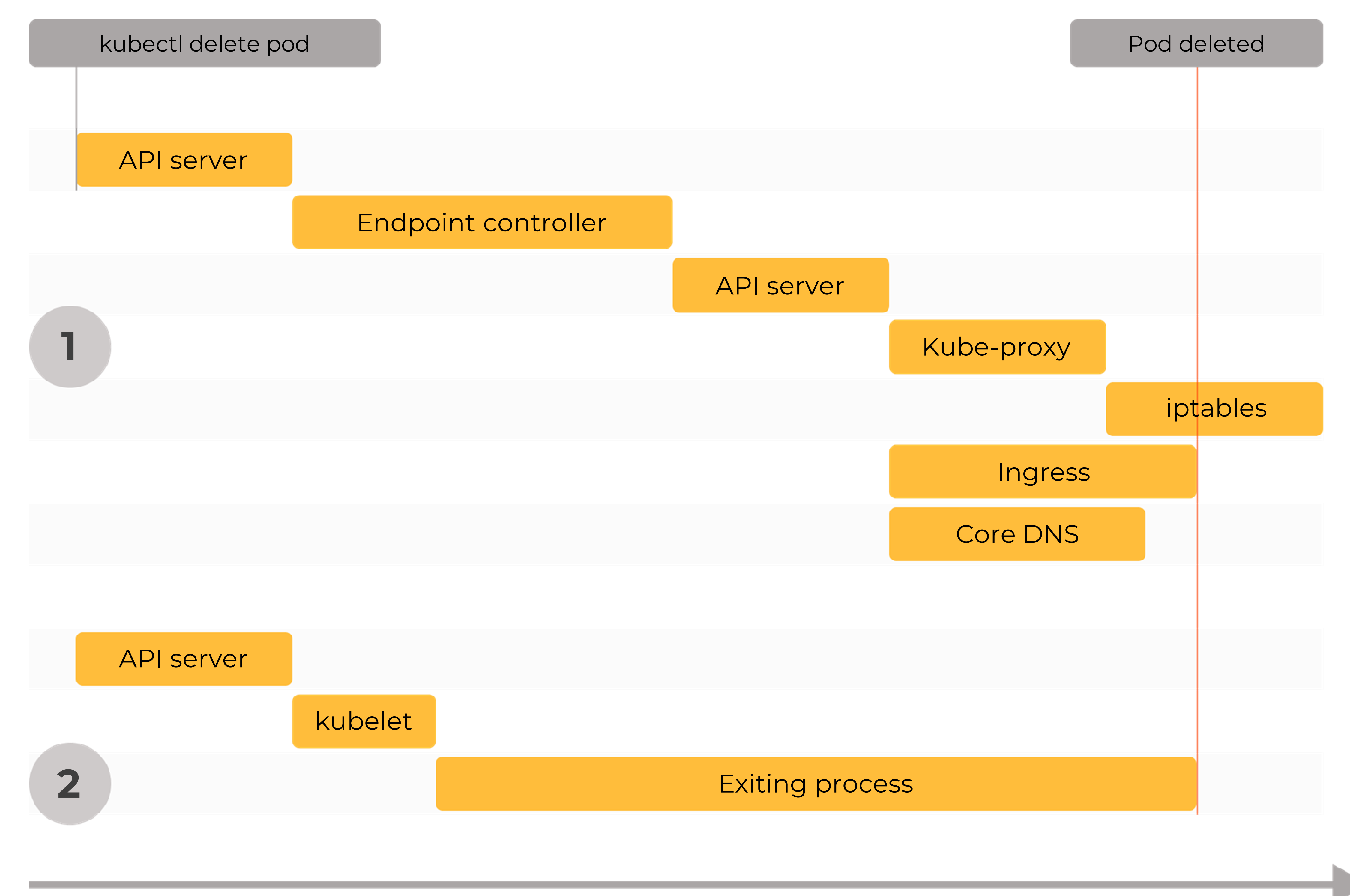


История из жизни №2:

- Приложение управляется через Deployment;
- У приложения 3 реплики;
- Стратегия обновления RollingUpdate;
- При выкате релиза кратковременно получаем 500-ые;
- Также бонусом получаем втык от начальства.



Почему так:


kubectl delete pod
Pod deleted
API server
Endpoint controller
API server
1
Kube-proxy
iptables
Ingress
Core DNS
API server
kubelet
2
Exiting process
Time

Почему так:

kubectl delete pod

Pod deleted

1

API server

Endpoint controller

API server

Kube-proxy

iptables

Ingress

Core DNS

SIGTERM

SIGKILL

2

API server

kubelet

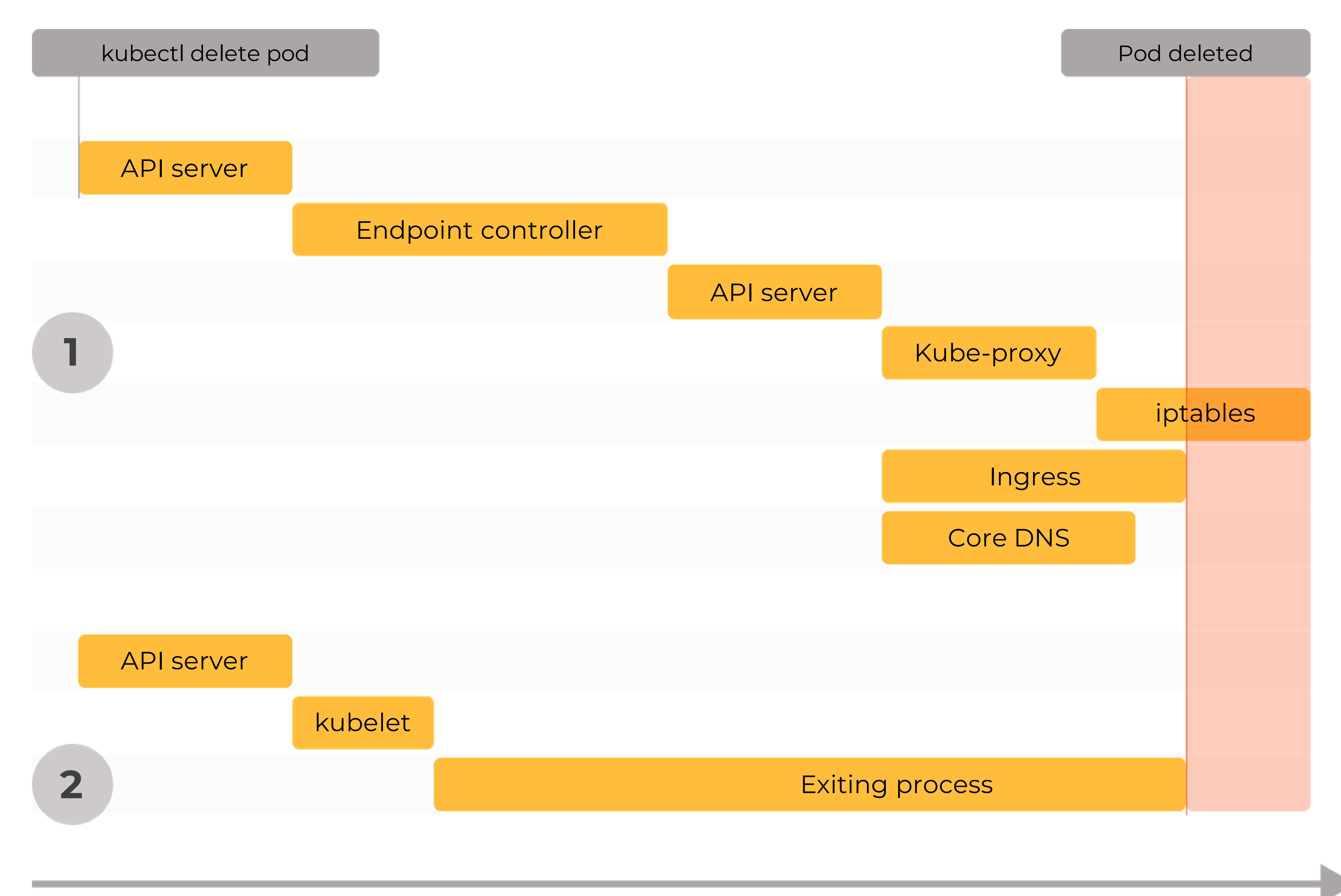
Exiting process

30 секунд

Time



Почему так:



kubectl delete pod
Pod deleted
API server
Endpoint controller
API server
1
Kube-proxy
iptables
Ingress
Core DNS
API server
kubelet
2
Exiting process
Time

Почему так:

kubectl delete pod

Pod deleted

1

API server

Endpoint controller

API server

Kube-proxy

iptables

Ingress

Core DNS

2

API server

kubelet

Exiting process

Time



Как решить вопрос:

- Убедиться, что приложение адекватно реагирует на SIGTERM;
- Подключить PreStop хуку;
- Добавить в нее «sleep 5» или, если не хватит, то 10 секунд.





```yaml
apiVersion: v1
kind: Pod
metadata:
  name: lifecycle-demo
spec:
  containers:
  - name: lifecycle-demo-container
    image: nginx
    lifecycle:
      preStop:
        exec:
          command: ["/bin/sh","-c", "sleep 5
&& /usr/sbin/nginx -s quit"]
```



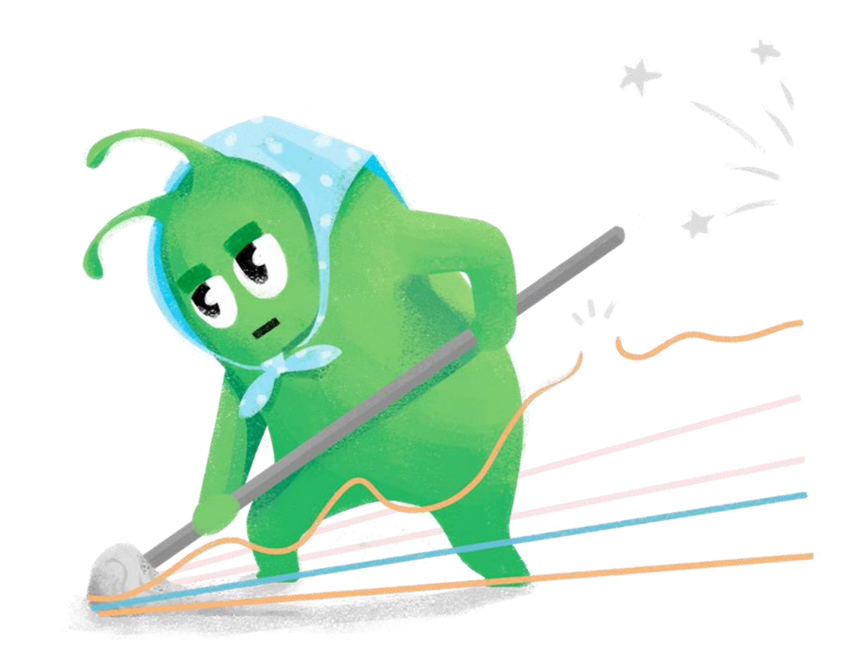


Как решить вопрос:

```yaml
image: {{ .Values.image }}:{{ .Values.imageTag}}
command:
  - /bin/sh
  - -c
  - bundle exec sidekiq -t 2400 -C config/sidekiq.yml --index 0 --environment
imagePullPolicy: IfNotPresent
name: sidekiq
lifecycle:
  preStop:
    exec:
      command:
        - /bin/sh
        - -c
        - kill -SIGTSTP
```




Как бороться с **человеческим фактором**?





История из жизни №3:

- Разработчики забывают проставить реквесты/лимиты;
- … ставят их некорректно;
- Как следствие – некорректный менеджмент ресурсов в кластере.





Limitrange

Устанавливает ресурсы для объектов кластера

Дефолтные

Максимальные

Минимальные

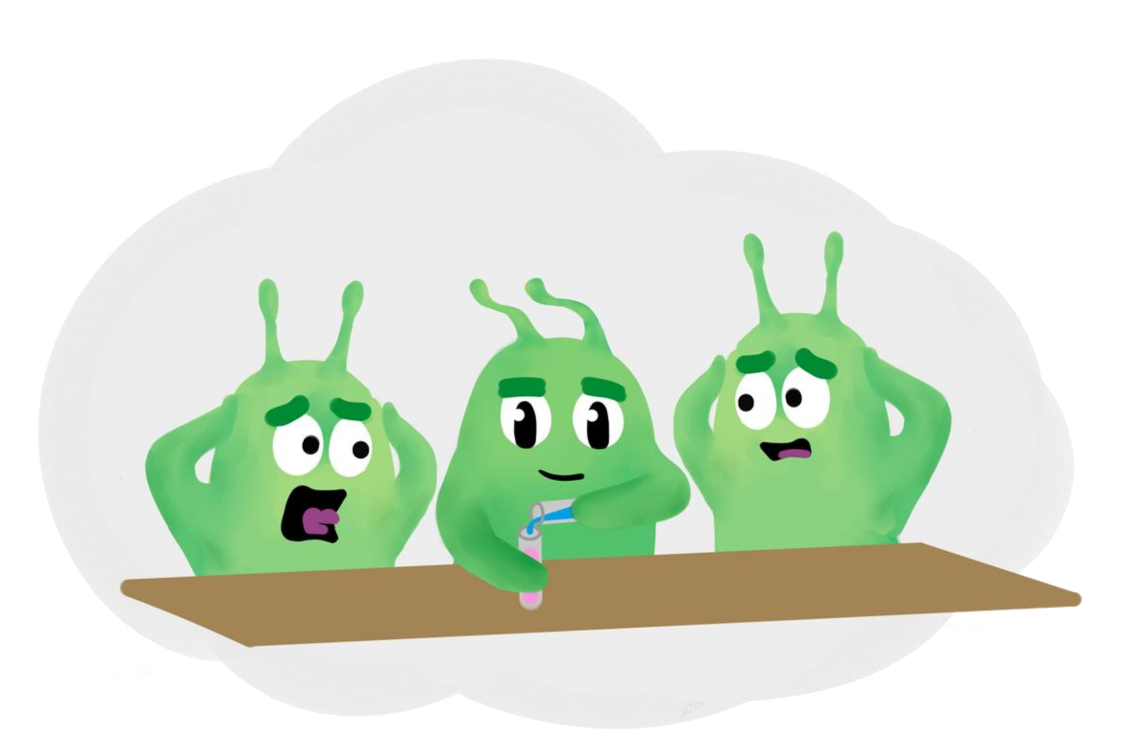
- Для контейнеров
- Для подов
- Для PVC

https://kubernetes.io/docs/concepts/policy/limit-range/



История из жизни №4:

- Разработчик ошибся с количеством реплик при скейлинге;
- Моргнул интернет, заела клавиша – и вот у нас 111 млн. реплик.





История из жизни №5:

- Настроили кластер, отдали необходимые доступы клиенту;
- Спустя полтора года в кластере мусорка.





Resourcequota

Устанавливает количество доступных ресурсов
и объектов для нэймспэйса в кластере

Реквесты

Лимиты

Сервисы

Поды

. . .

https://kubernetes.io/docs/concepts/policy/resource-quotas/



История из жизни №6:

- В одном кластере работали два окружения – prod и dev;
- В кластере случилась авария, и часть серверов стали недоступны;
- Поды начали активно эвакуироваться на рабочие ноды;
- Угадайте, поды какого окружения благополучно переехали, а какого не влезли до конца.

Priority Class

Устанавливает различные классы подов для шедулинга

В случае нехватки места для подов, наименее приоритетные поды будут удалены с нод

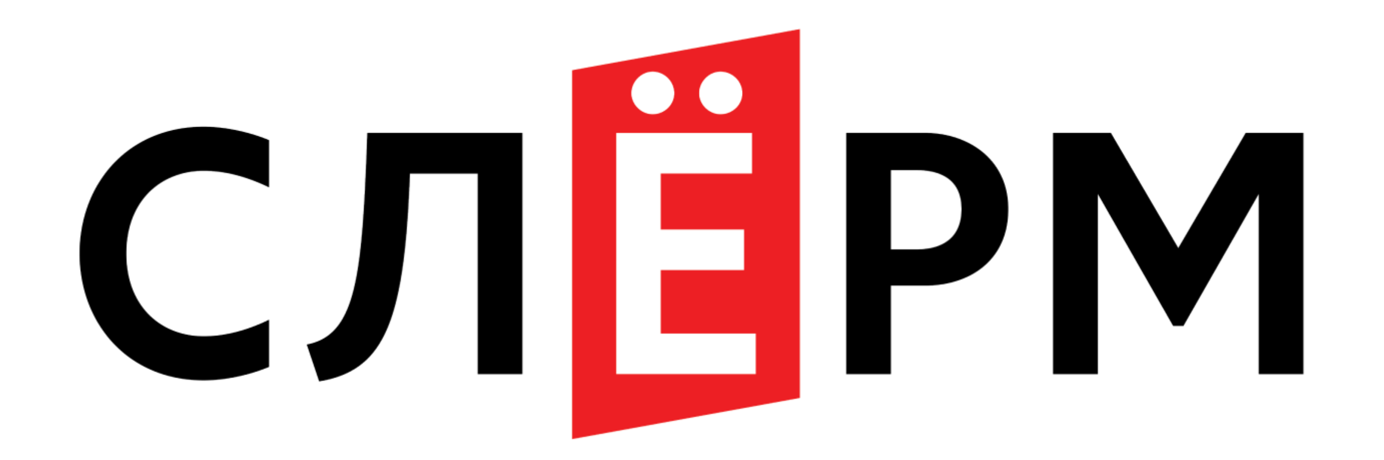
По приорити классам можно разделять ресурсы в квотах

https://kubernetes.io/docs/concepts/configuration/pod-priority-preemption/





ВСЁ !



— Учебный центр для IT-специалистов

— Обучили более 2000 инженеров

— В том числе целые компании и отделы

— Первый курс про Kubernetes в РФ

— Первыми стали Kubernetes Training Partners в РФ

https://slurm.io/

Telegram: @marsel91

